1 One- and Two-Sample Estimation Problems

1.1 Introduction

In previous chapters, we emphasized sampling properties of the sample mean and variance. The purpose of these presentations is to build a foundation that allows us to draw conclusions about the population parameters from experimental data.

1.2 Classical Methods of Estimation

A point estimate of some population parameter θ is a single value $\widehat{\theta}$ of a statistic $\widehat{\Theta}$. For example, the value $\overline{x}$ of the statistic $\overline{X}$, computed from a sample of size n, is a point estimate of the population parameter μ. Similarly, $\widehat{p} = x/n$ is a point estimate of the true proportion p for a binomial experiment.

An estimator is not expected to estimate the population parameter without error. We do not expect $\overline{X}$ to estimate μ exactly, but we certainly hope that it is not far off.

1.2.1 Unbiased Estimator

What are the desirable properties of a “good” decision function that would influence us to choose one estimator rather than another? Let $\widehat{\Theta}$ be an estimator whose value $\widehat{\theta}$ is a point estimate of some unknown population parameter θ. Certainly, we would like the sampling distribution of $\widehat{\Theta}$ to have a mean equal to the parameter estimated. An estimator possessing this property is said to be unbiased.

Definition 1 *A statistic $\widehat{\Theta}$ is said to be an unbiased estimator of the parameter θ if $\mu_{\widehat{\Theta}} = E(\widehat{\Theta}) = \theta$.*

Example 2 *Show that S^2 is an unbiased estimator of the parameter σ^2. Hint: $(X_i - \overline{X}) = (X_i - \mu) - (\overline{X} - \mu)$.*

1.2.2 Variance of a Point Estimator

If $\widehat{\Theta}_1$ and $\widehat{\Theta}_2$ are two unbiased estimators of the same population parameter θ, we want to choose the estimator whose sampling distribution has the smaller variance.

Hence, if $\sigma^2_{\widehat{\theta}_1} < \sigma^2_{\widehat{\theta}_2}$, we say that $\widehat{\Theta}_1$ is a more efficient estimator of θ than $\widehat{\Theta}_1$.

Definition 3 *If we consider all possible unbiased estimators of some parameter θ, the one with the smallest variance is called the most efficient estimator of θ.*

1.2.3 Interval Estimation

Even the most efficient unbiased estimator is unlikely to estimate the population parameter exactly. There is no reason we should expect a **point estimate** from a given sample to be exactly equal to the population parameter it is supposed to estimate. There are many situations in which it is preferable to determine an interval within which we would expect to find the value of the parameter. Such an interval is called an interval estimate. An interval estimate of a population parameter θ is an interval of the form $\widehat{\theta}_L < \theta < \widehat{\theta}_U$, where $\widehat{\theta}_l$ and $\widehat{\theta}_U$ depend on the value of the statistic $\widehat{\Theta}$ for a particular sample and also on the sampling distribution of $\widehat{\Theta}$.

2 Single Sample: Estimating the Mean

The sampling distribution of $\overline{X}$ is centered at μ, and in most applications the variance is smaller than that of any other estimators of μ. Thus, the sample mean $\overline{x}$ will be used as a point estimate for the population mean μ.

Let us now consider the interval estimate of μ. If our sample is selected from a normal population or, failing this, if n is sufficiently large, we can establish a confidence interval for μ by considering the sampling distribution of $\overline{X}$.

Definition 4 (Interval on μ, σ^2) *If $\overline{x}$ is the mean of a random sample of size n from a population with known variance σ^2, a* $100(1-\alpha)\%$ *confidence interval for μ is given by*

$$\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

Example 5 *The average zinc concentration recovered from a sample of measurements taken in 36 different locations in a river is found to be 2.6 grams per milliliter. Find the 95% and 99% confidence intervals for the mean zinc concentration in the river. Assume that the population standard deviation is 0.3 gram per milliliter.*

Solution 6 *The point estimate of μ is $\overline{x} = 2.6$. The z-value leaving an area of* 0.025 *to the right, and therefore an area of* 0.975 *to the left, is $z_{0.025} = 1.96$ (Table A.3). Hence, the* 95% *confidence interval is*

$$2.6 - (1.96)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (1.96)\left(\frac{0.3}{\sqrt{36}}\right)$$

which reduces to $2.50 < \mu < 2.70$. To find a 99% *confidence interval, we find the z-value leaving an area of* 0.005 *to the right and* 0.995 *to the left. From Table A.3* *again,* $z_{0.005} = 2.575$*, and the* 99% *confidence interval is*

$$2.6 - (2.575)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (2.575)\left(\frac{0.3}{\sqrt{36}}\right)$$

or simply

$$2.47 < \mu < 2.73.$$

The error in estimating μ by $\overline{x}$ is the absolute value of the difference between μ and $\overline{x}$, and we can be $100(1-\alpha)\%$ confident that this difference will not exceed $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

Theorem 7 *If $\overline{x}$ is used as an estimate of μ, we can be $100(1-\alpha)\%$ confident that the error will not exceed $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.*

Theorem 8 *If $\overline{x}$ is used as an estimate of μ, we can be $100(1-\alpha)\%$ confident that the error will not exceed a specified amount e when the sample size is*

$$n = \left(\frac{z_{\alpha/2}\sigma}{e}\right)^2$$

Example 9 *How large a sample is required if we want to be* 95% *confident that our estimate of* μ *in Example 5 is off by less than* 0.05*?*

Solution 10 *The population standard deviation is* $\sigma = 0.3$*. Then,*

$$n = \left[\frac{(1.96)(0.3)}{0.05}\right]^2 = 138.3.$$

Therefore, we can be 95% confident that a random sample of size 139 *will provide an estimate* $\overline{x}$ *differing from* μ *by an amount less than* 0.05*.*

The reader should recall learning in Chapter 3 that if we have a random sample from a normal distribution, then the random variable

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has a Student t-distribution with $n-1$ degrees of freedom. Here S is the sample standard deviation. In this situation, with σ unknown, T can be used to construct a confidence interval on μ.

Definition 11 (Interval on μ, σ^2) *If $\overline{x}$ and s are the mean and standard deviation of a random sample of size n from a normal population with unknown variance σ^2, a $100(1-\alpha)\%$ confidence interval for μ is*

$$\overline{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \overline{x} + t_{\alpha/2}\frac{s}{\sqrt{n}},$$

where $t_{\alpha/2}$ is the t-value with $\nu = n - 1$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

Example 12 *The contents of seven similar containers of sulfuric acid are 9.8, 10.2, 10.4, 9.8, 10.0, 10.2, 9.6 liters. Find a 95% confidence interval for the mean contents of all such containers, assuming an approximately normal distribution.*

Solution 13 *The sample mean and standard deviation for the given data are*

$$\overline{x} = 10.0 \text{ and } s = 0.283.$$

Using Table A.4, we find $t_{0.025} = 2.447$ *for* $v = 6$ *degrees of freedom. Hence, the* 95% *confidence interval for* μ *is*

$$10.0-(2.447)\left(\frac{0.283}{\sqrt{7}}\right) < \mu < 10.0+(2.447)\left(\frac{0.283}{\sqrt{7}}\right)$$

which reduces to $9.74 < \mu < 10.26$.

Concept of a Large-Sample Confidence Interval

Often statisticians recommend that even when normality cannot be assumed, σ is

unknown, and $n \geq 30$, s can replace σ and the confidence interval

$$\overline{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$$

may be used. This is often referred to as a large-sample confidence interval.

Example 14 *Scholastic Aptitude Test (SAT) mathematics scores of a random sample of* 500 *high school seniors in the state of Texas are collected, and the sample mean and standard deviation are found to be* 501 *and* 112, *respectively. Find a* 99% *confidence interval on the mean SAT mathematics score for seniors in the state of Texas.*

3 Standard Error of a Point Estimate

We indicated earlier that a measure of the quality of an unbiased estimator is its variance. The variance of $\overline{X}$ is

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$$

Thus, the standard deviation of $\overline{X}$, or standard error of $\overline{X}$, is $\sigma/\sqrt{n}$. Simply put, the standard error of an estimator is its standard deviation. For $\overline{X}$, the computed confidence limit

$$\overline{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \text{ is written } \overline{x} \pm z_{\alpha/2} \text{ s.e.}(\overline{x})$$

In the case where σ is unknown and sampling is from a normal distribution, s replaces σ and the estimated standard error $s/\sqrt{n}$ is involved. Thus, the confidence limits on μ are limit

$$\overline{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \text{ is written } \overline{x} \pm t_{\alpha/2} \text{ s.e.}(\overline{x})$$

4 Two Samples: Estimating the Difference between Two Means

Theorem 15 Confidence Interval for $\mu_1 - \mu_2$, σ_1^2 and σ_2^2 known
If $\overline{x}_1$ and $\overline{x}_2$ are means of independent random samples of sizes n_1 and n_2 from populations with known variances σ_1^2 and σ_2^2, respectively, a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$(\overline{x}_1 - \overline{x}_2) - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\overline{x}_1 - \overline{x}_2) + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},$$

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

Example 16 *A study was conducted in which two types of engines, A and B, were compared. Gas mileage, in miles per gallon, was measured. Fifty experiments were conducted using engine type A and 75 experiments were* *done with engine type B. The gasoline used and other conditions were held constant. The average gas mileage was* 36 *miles per gallon for engine A and* 42 *miles per gallon for engine B. Find a* 96% *confidence interval on $\mu_B-\ \mu_A$, where μ_A and μ_B are population mean gas mileages for engines A and B, respectively. Assume that the population standard deviations are* 6 *and* 8 *for engines and B, respectively.*

Solution 17 *The point estimate of $\mu_B-\ \mu_A$ is $\overline{x}_B - \overline{x}_A$ $= 42 - 36 = 6$. Using $\alpha = 0.04$, we find $z_{0.02} = 2.05$ from Table A.3. Hence, with substitution in the formula above, the* 96% *confidence interval is*

$$6 - 2.05\sqrt{\frac{64}{75} + \frac{36}{50}} < \mu_B - \mu_A < 6 + 2.05\sqrt{\frac{64}{75} + \frac{36}{50}}$$

or simply $3.43 < \mu_B - \mu_A < 8.57$.

Variances Unknown but Equal

Consider the case where σ_1^2 and σ_2^2 are unknown and $\sigma_1^2 = \sigma_1^2\,(= \sigma^2)$. A point estimate of the unknown common variance σ^2 can be obtained by pooling the sample variances. Denoting the pooled estimator by S_p^2, we have the following.

Definition 18 (of Variance) $S_p^2 = \frac{(n_1-1)S_1^2+(n_1-1)S_2^2}{(n_1+n_2-1)}$

Theorem 19 Confidence Interval for $\mu_1 - \mu_2$, $\sigma_1^2 = \sigma_2^2$ but Both Uknown
If $\overline{x}_1$ and $\overline{x}_2$ are means of independent random samples of sizes n_1 and n_2, respectively, from approximately normal populations with unknown but equal variances, a $100(1-\alpha)\%$ *confidence interval for $\mu_1 - \mu_2$ is given by*

$$(\overline{x}_1-\overline{x}_2)-t_{\frac{\alpha}{2}}s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}<\mu_1-\mu_2<(\overline{x}_1-\overline{x}_2)+t_{\frac{\alpha}{2}}s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},$$

where where s_p is the pooled estimate of the population standard deviation and $t_{\alpha/2}$ is the t-value with $\nu = n_1 + n_2 - 2$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

Example 20 *Two independent sampling stations, statoin 1 and station 2, were chosen for a study on pollution. For 12 monthly samples collected at station 1, the species diversity index had a mean value $\overline{x}_1 = 3.11$ and a standard deviation $s_1 = 0.771$, while 10 monthly samples collected at the station 2 had a mean index value $\overline{x}_2 = 2.04$ and a standard deviation $s_2 = 0.448$. Find a 90% confidence interval for the difference between the population means for the two locations, assuming that the populations are approximately normally distributed with equal variances.*

Solution 21 *Let μ_1 and μ_2 represent the population means, respectively, for the species diversity indices at the downstream and upstream stations. We wish to find a 90%*

confidence interval for $\mu_1 - \mu_2$. Our point estimate of $\mu_1 - \mu_2$ is

$$\overline{x}_1 - \overline{x}_2 = 3.11 - 2.04 = 1.07.$$

The pooled estimate, s_p^2, of the common variance, σ^2, is

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_1 - 1)s_2^2}{(n_1 + n_2 - 1)} = \frac{(11)(0.7712)+(9)(0.4482)}{12+10-2} = 0.417.$$

Taking the square root, we obtain $s_p = 0.646$. Using $\alpha = 0.1$, we find in Table A.4 that $t_{0.05} = 1.725$ for $\nu = n_1 + n_2 - 2 = 20$ degrees of freedom. Therefore, the 90% *confidence interval for $\mu_1 - \mu_2$ is*

$$\begin{aligned} 1.07+1.725(0.646)\sqrt{\tfrac{1}{12}+\tfrac{1}{10}} &< \mu_1-\mu_2 \\ &< 1.07+1.725(0.646)\sqrt{\tfrac{1}{12}+\tfrac{1}{10}} \end{aligned}$$

which simplifies to $0.593 < \mu_1 - \mu_2 < 1.547$.

5 Paired Observations

Now we shall consider estimation procedures for the difference of two means when the samples are not independent and the variances of the two populations are not necessarily equal. The situation considered here deals with a very special experimental condition, namely that of paired observations. For example, if we run a test on a new diet using 15 individuals, the weights before and after going on the diet form the information for our two samples. The two populations are "before" and "after," and the experimental unit is the individual. Obviously, the observations in a pair have something in common. To determine if the diet is effective, we consider the differences $d_1, d_2, ..., d_n$ in the paired observations. These differences are the values of a random sample $D_1, D_2, ..., D_n$ from a population of differences that we shall assume to be normally distributed with mean $\mu_D = \mu_1 - \mu_2$ and variance σ_D^2. We estimate σ_D^2 by σ_d^2, the variance of the differences that constitute our sample. The point estimator of μ_D is given by $\overline{D}$.

Theorem 22 Confidence Interval for $\mu_D = \mu_1 - \mu_2$, for Paired Observations
If $\overline{d}$ and s_d are the mean and standard deviation, respectively, of the normally distributed differences of n random pairs of measurements, a $100(1-\alpha)\%$ confidence interval for $\mu_D = \mu_1 - \mu_2$ is

$$\overline{d} - t_{\alpha/2}\frac{s_d}{\sqrt{n}} < \mu < \overline{d} + t_{\alpha/2}\frac{s_d}{\sqrt{n}},$$

where $t_{\alpha/2}$ is the t-value with $\nu = n - 1$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

Example 23 *A study published in Chemosphere reported the levels of the dioxin TCDD of 10 Massachusetts Vietnam veterans who were possibly exposed to Agent Orange. The TCDD levels in plasma and in fat tissue are listed in Table 1. Find a 95% confidence interval for $\mu_1 - \mu_2$, where μ_1 and μ_2 represent the true mean TCDD levels in plasma and in fat tissue, respectively. Assume the distribution of the differences to be approximately normal.*

Veteran	TCDD levels in Plasma	TCDD levels in Fat Tissue	d_i
1	2.5	4.9	-2.4
2	3.1	5.9	-2.8
3	2.1	4.4	-2.3
4	3.5	6.9	-3.4
5	3.1	7.0	-3.9
6	1.8	4.2	-2.4
7	6.0	10.0	-4.0
8	3.0	5.5	-2.5
9	36.0	41.0	-5.0
10	4.7	4.4	0.3

Solution 24 *The point estimate of* μ_D *is* $\overline{d} = -2.84$*. The standard deviation,* s_d*, of the sample differences is* 1.42*. Using* $\alpha = 0.05$*, we find in Table A.4 that* $t_{0.025} =$ 2.262 *for* $\nu = n - 1 = 9$ *degrees of freedom. Therefore, the* 95% *confidence interval is*

$$-2.84-(2.262)\left(\frac{1.42}{\sqrt{10}}\right) < \mu_D < -2.84+(2.262)\left(\frac{1.42}{\sqrt{10}}\right)$$

or simply $-3.85 < \mu_D < -1.82$.

6 Single Sample: Estimating a Proportion

A point estimator of the proportion p in a binomial experiment is given by the

statistic $\widehat{P} = X/n$, where X represents the number of successes in n trials. Therefore,

Definition 25 *the sample proportion $\widehat{p} = x/n$ will be used as the point estimate of the*

parameter p.

Theorem 26 (Large-Sample Confidence Intervals for p) *If $\widehat{p}$ is the proportion of successes in a random sample of size n and $\widehat{q} = 1- \widehat{p}$, an approximate* $100(1-\alpha)\%$

confidence interval, for the binomial parameter p is given by

$$\widehat{p} - z_{\alpha/2}\sqrt{\frac{\widehat{p}\widehat{q}}{n}} < p < \widehat{p} + z_{\alpha/2}\sqrt{\frac{\widehat{p}\widehat{q}}{n}}$$

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

Example 27 *In a random sample of $n = 500$ families owning television sets in the city of Hamilton, Canada, it is found that $x = 340$ subscribe to HBO. Find a 95% confidence interval for the actual proportion of families with television sets in this city that subscribe to HBO.*

Solution 28 *The point estimate of p is $\widehat{p} = 340/500 = 0.68$. Using Table A.3, we find that $z_{0.025} = 1.96$. Therefore, the 95% confidence interval for p is*

$$0.68 - 1.96\sqrt{\frac{(0.68)(0.32)}{500}} < p < 0.68 + 1.96\sqrt{\frac{(0.68)(0.32)}{500}}$$

which simplifies to $0.6391 < p < 0.7209$.

Theorem 29 *If $\widehat{p}$ is used as an estimate of p, we can be* $100(1-\alpha)\%$ *confident that the error will not exceed* $z_{\alpha/2}\sqrt{\frac{\widehat{pq}}{n}}$.

Choice of Sample Size

Let us now determine how large a sample is necessary to ensure that the error in estimating p will be less than a specified amount e. By Theorem 23, we must choose n such that $z_{\alpha/2}\sqrt{\frac{\widehat{pq}}{n}} = e$.

Theorem 30 *If $\widehat{p}$ is used as an estimate of p, we can be* $100(1-\alpha)\%$ *confident that the error will be less than a specified amount e when the sample size is approximately*

$$n = \frac{z_{\alpha/2}^{2}\widehat{p}\widehat{q}}{e^{2}}$$

Example 31 *How large a sample is required if we want to be 95% confident that our estimate of p in Example 21 is within* 0.02 *of the true value?*

Solution 32 *Let us treat the* 500 *families as a preliminary sample, providing an estimate* $\widehat{p} = 0.68$. *Then,*

$$n = \frac{(1.96)^2(0.68)(0.32)}{0.02^2} = 2089.8 \approx 2090$$

Occasionally, it will be impractical to obtain an estimate of p to be used for determining the sample size for a specified degree of confidence. If this happens, we use the following theorem.

Theorem 33 *If* $\widehat{p}$ *is used as an estimate of* p, *we can be* $100(1-\alpha)\%$ *confident that the error will not exceed than a specified amount* e *when the sample size is approximately*

$$n = \frac{z^2_{\alpha/2}}{4e^2}$$

Example 34 *How large a sample is required if we want to be at least* 95% *confident that our estimate of* p *in Example 21 is within* 0.02 *of the true value?*

Solution 35 *Let assume that no preliminary sample has been taken to provide an estimate of p. Consequently, we can be at least* 95% *confident that our sample proportion will not differ from the true proportion by more than* 0.02 *if we choose a sample of size*

$$n = \frac{(1.96)^2}{4(0.02)^2} = 2401$$

Comparing the results of Examples 28 and 29, we see that information concerning p, *provided by a preliminary sample or from experience, enables us to choose a smaller sample while maintaining our required degree of accuracy.*

7 Two Samples: Estimating the Difference between Two Proportions

Consider the problem where we wish to estimate the difference between two binomial

parameters p_1 and p_2. For example, p_1 might be the proportion of smokers

with lung cancer and p_2 the proportion of nonsmokers with lung cancer, and the

problem is to estimate the difference between these two proportions.

Theorem 36 Large-Sample Confidence Interval for $p_1 - p_2$
If $\widehat{p}_1$ *and* $\widehat{p}_2$ *are the proportions of successes in random samples of sizes* n_1 *and* n_2*, respectively,* $\widehat{q}_1 = 1-\widehat{p}_1$*, and*

$\widehat{q}_2 = 1 - \widehat{p}_2$, *an approximate* $100(1-\alpha)\%$ *confidence interval for the difference of two binomial parameters,* $p_1 - p_2$, *is given by*

$$(\widehat{p}_1 - \widehat{p}_2)\text{-}z_{\frac{\alpha}{2}}\sqrt{\frac{\widehat{p}_1\widehat{q}_1}{n_1} + \frac{\widehat{p}_2\widehat{q}_2}{n_1}} < p_1 - p_2 < (\widehat{p}_1 - \widehat{p}_2) + z_{\frac{\alpha}{2}}\sqrt{\frac{\widehat{p}_1\widehat{q}_1}{n_1} + \frac{\widehat{p}_2\widehat{q}_2}{n_1}}$$

Example 37 *A certain change in a process for manufacturing component parts is being considered. Samples are taken under both the existing and the new process so as to determine if the new process results in an improvement. If* 75 *of* 1500 *items from the existing process are found to be defective and* 80 *of* 2000 *items from the new process are found to be defective, find a* 90% *confidence interval for the true difference in the proportion of defectives between the existing and the new process.*

Solution 38 *Let* p_1 *and* p_2 *be the true proportions of defectives for the existing and new processes, respectively. Hence,* $\widehat{p}_1 = 75/1500 = 0.05$ *and* $\widehat{p}_2 = 80/2000 = 0.04$, *and the point estimate of* $p_1 - p_2$ *is*

$$\widehat{p}_1 - \widehat{p}_2 = 0.05 - 0.04 = 0.01$$

Using Table A.3, we find $z_{0.05} = 1.645$*. Therefore, substituting into the formula, with*

$$1.645\sqrt{\frac{(0.05)(0.95)}{1500} + \frac{(0.04)(0.96)}{2000}} = 0.0117,$$

we find the 90% *confidence interval to be* $-0.0017 < p_1 - p_2 < 0.0217$*.*

8 Single Sample: Estimating the Variance

If a sample of size n is drawn from a normal population with variance σ^2 and the sample variance s^2 is computed, we obtain a value of the statistic S^2. This computed sample variance is used as a point estimate of σ^2. Hence, the statistic S^2 is called an estimator of σ^2. An interval estimate of σ^2 can be established by using the statistic

$$X = \frac{(n-1)S^2}{\sigma^2}$$

the statistic X has a chi-squared distribution with $n-1$ degrees of freedom when samples are chosen from a normal population.

Theorem 39 (Confidence Interval for σ^2) *If s^2 is the variance of a random sample of size n from a normal population, a $100(1-\alpha)\%$ confidence interval for σ^2 is*

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

where $\chi^2_{\alpha/2}$ *and* $\chi^2_{1-\alpha/2}$ *are* χ^2*-values with* $\nu = n - 1$ *degrees of freedom, leaving areas of* $\alpha/2$ *and* $1-\ \alpha/2$*, respectively, to the right.*

An approximate $100(1 - \alpha)\%$ confidence interval for σ is obtained by taking the square root of each endpoint of the interval for σ^2.

Example 40 *The following are the weights, in decagrams, of 10 packages of grass seed distributed by a certain company:* $46.4, 46.1, 45.8, 47.0, 46.1, 45.9, 45.8, 46.9, 45.2, 46.0$*. Find a* 95% *confidence interval for the variance of the weights of all such packages of grass seed distributed by this company, assuming a normal population.*

Solution 41 *First we find* $s^2 = 0.286$*. To obtain a* 95% *confidence interval, we choose* $\alpha = 0.05$*. Then, using Table A.5 with* $\nu = 9$ *degrees of freedom, we find* $\chi^2_{.025} = 19.023$ *and* $\chi^2_{.975} = 2.700$*. Therefore, the* 95% *confidence interval for* σ^2 *is*

$$\frac{(9)(0.286)}{19.023} < \sigma^2 < \frac{(9)(0.286)}{2.700}$$

or simply $0.135 < \sigma^2 < 0.953$*.*

9 Two Samples: Estimating the Ratio of Two Variances

A point estimate of the ratio of two population variances σ_1^2/σ_2^2 is given by the ratios s_1^2/s_2^2 of the sample variances. Hence, the statistic S_1^2/S_2^2 is called an estimator of σ_1^2/σ_2^2 . If σ_1^2 andσ_2^2 are the variances of normal populations, we can establish an interval estimate of σ_1^2/σ_2^2 by using the statistic

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

According to Theorem 25 of chapter 3, the random variable F has an F-distribution with $\nu_1 = n_1 - 1$ and ν_2 $= n_2 - 1$ degrees of freedom.

Theorem 42 (Confidence Interval for σ_1^2/σ_2^2) *If s_1^2 and s_2^2 are the variances of independent samples of sizes n_1*

and n_2*, respectively,from normal populations, then a* $100(1-\alpha)\%$ *confidence interval for* σ_1^2/σ_2^2 *is*

$$\frac{s_1^2}{s_2^2}\frac{1}{f_{\alpha/2}(\nu_1,\nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2}f_{\alpha/2}(\nu_2,\nu_1)$$

where $f_{\alpha/2}(\nu_1,\nu_2)$ *is an* f*-value with* $\nu_1 = n_1 - 1$ *and* $\nu_2 = n_2 - 1$ *degrees of freedom, leaving an area of* $\alpha/2$ *to the right, and* $f_{\alpha/2}(\nu_2,\nu_1)$ *is a similar* f*-value with* $\nu_2 = n_2 - 1$ *and* $\nu_1 = n_1 - 1$ *degrees of freedom.*

an approximate $100(1-\alpha)\%$ confidence interval for σ_1/σ_2 is obtained by taking the square root of each endpoint of the interval for σ_1^2/σ_2^2.

Example 43 *A study was conducted to estimate the difference in the amounts of the chemical orthophosphorus measured at two different stations. Fifteen samples were collected from station 1, and* 12 *samples were obtained from station 2. The* 15 *samples from station 1 had an average orthophosphorus content of* 3.84 *milligrams per* *liter and a standard deviation of* 3.07 *milligrams per liter, while the* 12 *samples from station 2 had an average content of* 1.49 *milligrams per liter and a standard deviation of* 0.80 *milligram per liter. Determine a 98% confidence interval for* σ_1^2/σ_2^2 *and for* σ_1/σ_2*, where* σ_1^2 *and* σ_2^2 *are the variances of the populations of orthophosphorus contents at station* 1 *and station* 2*, respectively.*

Solution 44 *We have* $n = 15$, $n_2 = 12$, $s_1 = 3.07$, *and* $s_2 = 0.80$. *For a* 98% *confidence interval,* $\alpha = 0.02$. *Interpolating in Table A.6, we find* $f_{0.01}(14, 11) \approx 4.30$ *and* $f_{0.01}(11, 14) \approx 3.87$. *Therefore, the* 98% *confidence interval for* σ_1^2/σ_2^2 *is*

$$\left(\frac{3.07^2}{0.80^2}\right)\left(\frac{1}{4.30}\right) < \frac{\sigma_1^2}{\sigma_2^2} < \left(\frac{3.07^2}{0.80^2}\right)(3.87),$$

which simplifies to $3.425 < \frac{\sigma_1^2}{\sigma_2^2} < 56.991$. *Taking square roots of the confidence limits, we find that a* 98% *confidence interval for* σ_1/σ_2 *is*

$$1.851 < \frac{\sigma_1}{\sigma_2} < 7.549.$$

Since this interval does not allow for the possibility of σ_1/σ_2 being equal to 1, we were correct in assuming that $\sigma_1 \neq \sigma_2$ (and $\sigma_1^2 \neq \sigma_2^2$).